

# 物語としての海馬体計算モデルの提案

## — 全脳アーキテクチャを構築するために —

### First story of hippocampal computing for developing whole brain architectures

山川 宏<sup>\*1</sup>  
Hiroshi YAMAKAWA

<sup>\*1</sup> 株式会社ドワンゴ ドワンゴ人工知能研究所  
DWANGO Co., ltd Dwango Artificial Intelligence Laboratory

The whole brain architecture (WBA) researches are aiming to build a general-purpose cognitive architecture. Computational models, which stand on neuroscientific facts, are necessary for these researches. For building computational model, iterative refining process of story is necessary. A story is a hypothesis network for converging piecemeal knowledge of neuroscience, and explaining expected cognitive function. For start up discussion, I propose an initial story about the computational model of the hippocampus and entorhinal cortex, in this paper.

## 1. はじめに

全脳アーキテクチャ[全脳]は汎用人工知能を構築するための一つの研究アプローチである。技術的目標は、脳の計算機能を再構成できるように、できるだけ高い抽象度で脳の器官に対応する機械学習装置を当てはめることで、一種のパズルを完成させて、人間並みかそれ以上の能力を持つ汎用の知能機械を作ることである。なお機械学習の進展を受けて、脳の構築に必要なパズルのピース(機械学習装置)は、すでに粗方揃っていると仮定している。一方で、神経科学の進展が著しいとはいえ、脳の神経システムを同定することで、そのまま知能を再現できるまでには、さらに数十年を要するであろう(全脳エミュレーション)。

パズルを解く主な方略は二つあり、ひとつは周辺部から確実に作り上げてゆく方法で、もう一つはオブジェクト等の意味のある繋がり(パズル中の車の図柄とか)を手掛かりに塊を作る方法である。これを自然科学における研究アプローチとして見ると、前者は再現可能な実験を通じてより客観性の高い事実を積み上げる物理分野を中心とした手堅いアプローチであるが、現状ではこの方略だけで壮大な脳のパズルを完成することは難しい。一方で、後者は歴史学の研究などにおいて再現できない証拠を繋ぎあわせて物語を作るアプローチに似ている。ここで物語は、論理的に矛盾のない仮説の集合体である[野家 05]。

神経科学的な知見が乏しかった 20 世紀においては、様々な研究者が思いのままに計算モデルを語ったとしても、それを神経科学的な知見から棄却することは難しかった。しかし、断片的であるとはいえ、現在の神経科学においては実験知見が急増してきた。よって、それらの知見に対して概ね整合しうる、有機的に結合された仮説集合としての物語は強く限定される、残された少数の物語は有力な候補となりうる。そこで全脳アーキテクチャを構築するための有力な研究アプローチは、こうした物語としての脳の計算モデルを大きな仮説集合として仮置きし、それを洗練・棄却するというプロセスに基づく。

### 1.1 物語作成の方法

物語の理想的な構築方法は、知り得る、神経科学、認知科学、人工知能、発達科学、ロボット工学などといった関連書分野のあらゆる知見からみて、もっとも棄却されにくい物語を生成することである。しかしこれらの膨大な知識を特定の個人が把握する

ことはできない。よって一つの有力な選択肢は、関連する知識を全て計算機が扱える形に加工した上で、それらを組み合わせることで制約充足問題を解くことであるが、現状ではそれは難しい。

そこで以下のプロセスの如く、複数の物語が語られることを通じ、脳の計算機能を実現するモデルへの到達を目指す。

- Step 1. 諸分野の重要な知見等を列挙する
- Step 2. 上記のうちの幾つかを組み合わせ何らかの物語 A を構築する(恣意性は可能な範囲で減らす)
- Step 3. 考慮するあらたな知見を選択する
  - ・ 知見の追加で A が改良されて A' になったら Step 3 へ
  - ・ 新たな知見により物語 A が棄却されたら Step 2 に戻る

こうした物語は複数存在し、通常はその物語を主導する研究者を軸として上記サイクルを繰り返すことになるだろう。

次章以降で、海馬-嗅内皮質の計算モデルについての物語を提案する。物語を構成する仮説要素の多くは、神経科学において、必ずしも目新しくはない。本稿の目的は、想定しうる仮説をつなぎ合わせたネットワーク化し物語を紡ぎだすことである。

## 2. 海馬-嗅内皮質の神経回路と空間選択細胞

海馬-嗅内皮質は、近年、齧歯類の動物実験を通じて神経科学知見が急速に蓄積し[北西 15]、改めて詳細なモデル化を行う準備が整いつつある。背景には、遺伝子改変技術や多数の神経細胞活動を行動中の動物で測定する技術の進展がある。

図 1 を用いて本稿で扱う主要な神経回路を説明する。図中の各丸は、ある部分器官における興奮性神経細胞を代表して図示し、各丸に対して、数十万個規模の神経細胞が並列に存在する。図示していない抑制性神経細胞が部分器官毎に局在することで興奮性細胞のスパースな神経活動を支えている。

### 2.1 トリシナプス回路と CA3 の再帰結合神経回路

海馬内の基本的な興奮伝達経路は、嗅内皮質から海馬歯状回への入力、CA3 領域を経て CA1 領域へ至り、再び皮質へ出力するループ構造である。この経路は貫通路、苔状線維、Schaffer 側枝による3つのシナプスを含むことからトリシナプス回路と呼ばれる。なお EC 内の深層から III 層へ、さらに II 層への伝達する経路を含む(図中赤線で示した)。行動中のマウス等においては、Medial septum から生成されたシータ波が、EC2 からこのループ回路に入力されて伝播する。そして CA3 は特

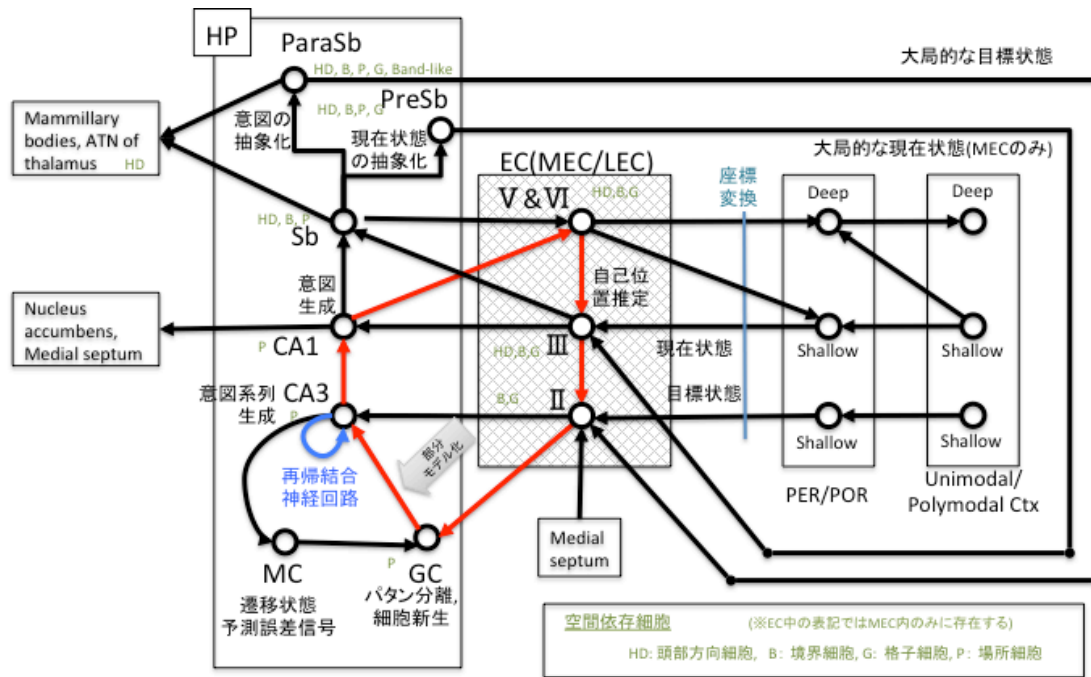


図1: 海馬-嗅内皮質の興奮性神経細胞による主要な神経回路

HP: Hippocampus (海馬), EC: Entorhinal Cortex(嗅内皮質)(MEC は内側, LEC は外側), GC: Granule Cell (顆粒細胞), MC: Mossy Cell (苔状細胞), Sb: Subiculum (海馬台), ParaSb: Parasubiculum (傍海馬台), PreSb: Presubiculum (前海馬台), Perirhinal Cortex (嗅周囲皮質), Postrhinal Cortex (嗅後部皮質), Mammillary bodies (乳頭体), Anterior nuclei of thalamus (視床前核), Nucleus accumbens (側坐核), Medial septum (内側中隔核), 赤い矢印でトリナプス回路を示した。

微的な再帰結合を持っており後述するシータ位相歳差現象は CA3 を起点としてトリナプス回路を伝播すると想定される。つまり海馬-嗅内皮質回路は、基本的に二段階ループ回路である。

### 3. What 経路と Where 経路

外界の認識において、「何(What)が何処(Where)にあるか?」という問いは重要であり、新皮質と海馬への接続部分で二つに情報は独立にまとめられている。What は新皮質から PER および LEC を通じて海馬に至る経路に、Where は、新皮質から POR および MEC を通じて海馬に至る経路に表現される。そして海馬内において、Where と What の情報が統合されている。

Where の情報は自分の移動に応じて、外界について一貫した変化(座標変換)を生じさせる必要があるのに対して、What の情報はより個別的であるためにこうした違いが有効であると考えられる。こうした考察と一致するが、空間のメトリックを担うとされる格子細胞は、Where を表現する、MEC に存在する。

#### 3.1 空間選択細胞

海馬において齧歯類を用いた実験が進展してきた要因は、動物が迷路などの中で課題を遂行する際に、空間情報(where 経路)に依存した活動を示す神経細胞が発見されたことが大きく、今回のモデル提案もそれらの知見に大きく依存している。これらは環境中心座標系で構成されており認知地図とも呼ばれる。

発達の順に見れば、動物の頭の向きに反応して発火する頭部方向細胞(HD)が PreSb, ParaSb MEC, 背側被蓋核, 乳頭体, 視床前核にまず形成される。次に動物から一定の方向かつ一定の距離に壁が存在するときに活動する境界細胞は, Sb, ParaSb, PreSb, MEC が発達する。空間中の特定の場所に反応する場所細胞(P)は GC, CA3, CA1, S, ParaSb, PreSb において経験に応じて段階的に形成される。これに対し EC 内では典型

的には二次元平面上を埋め尽くす格子点で発火する格子細胞(G)がある段階で一気に発達する。頭部方向細胞や境界細胞は前処理されたセンシング情報であるのに対し、場所細胞と格子細胞は認知地図を表現するものである。ただし場所細胞の性質は脳部位(神経回路)毎に少しずつ性質は異なる。

動物が広いフィールドで活動する場合には複数地点で場所細胞が活動することも知られていることから、格子細胞も場所細胞も集団によって特定の場所を指定できるという意味では同様である。しかし場所細胞の集団により指定できる特定の場所を表現することでランドマークや視覚刺激との対応付けた状態遷移モデルが利用できるようになるのだろう。一方、MEC 上の格子細胞は Postrhinal Cortex との座標変換に向くと思われる。

### 4. 海馬 CA3: 状態遷移モデルと意図系列生成

エージェントとしての動物の脳は、状態の系列として得られる経験に基づいて、自身のおかれた外界の環境を把握し、その上で個体や種として存続できるような知的振る舞いを行う。人工知能分野において知的エージェントが行動を生成する主要な手段は、効用志向, ゴール志向の二種類がある。そして、魚類から人間に至る高等な脊椎動物の脳では、これらが組み合わされて利用されていると考えるのが妥当であろう。

#### 4.1 状態遷移モデルとしての海馬 CA3

効用志向はエージェント自身にとっての状態の望ましさを評価し、それを高めるよう行動を生成する。これにより生物が個体や種として存続するために必要な固定的な価値を直接的に実装できる。この能力を格段に高める強化学習は、望ましい状態に達する以前の状態に対しても段階的に効用を付加する学習技術であり、長期的な視野で行動が行うことが可能になる。脳においては脳基底核を中心に実装されていることが定説であ



るが、経験で得られた状態のみに評価を与える強化学習のみでは学習効率が悪く現実的な性能を得られない<sup>1</sup>。

これに対してゴール志向のエージェントは、状態の遷移を表す状態遷移モデルを利用して、現在の状態から目的の状態にたどり着くための行動をプランニング(計画)する機能である。動物の脳では状態遷移モデルも経験を通じて獲得するであろうが、これは全ての経験を利用できる教師なし学習となり効率が良い。

脳は何れかの部分に、状態遷移モデルを実装していると思われる。再帰結合神経回路を持つ海馬 CA3 は状態遷移モデルとしてふさわしいが、伝統的には連想記憶モデルとのみかたが多かった。しかし次の理由により状態遷移モデルであると考える。一つ目に広く同一サイクル下で同期動作するリカレント神経回路は脳内には CA3 以外に存在しない。二つ目に CA3 の結合はリカレントであるため時系列情報の表現に適している。三つ目は意図系列の生成能力を担うからである(次節参照)。

よって本稿では、海馬 CA3 を状態遷移モデルとみなす以下を仮定し、以降はこの仮説の下で議論を進める。

**海馬 CA3 状態遷移モデル仮説:** 少なくともサブ秒から数秒の時間オーダにおける脳全体の状態遷移モデルは海馬 CA3 の再帰結合な神経回路により実現される。

## 4.2 意図系列を生成する CA3

意図(Intention)は BDI アーキテクチャで扱われる、現在状態に対する信念(Brief)と望ましい状態 D(Desire)に対置される、実現可能と信ずる望ましい状態である。ここでは「望ましい予測状態」と考える。そして意図系列とは、複数の意図が連なった系列であるとする。齧歯類の海馬 CA3 では遅いガンマ波(40Hz 程度)<sup>2</sup>で意図系列が SWR とシータ位相歳差現象として現れる。

動物が迷路上で停止時に発生する SWR (Sharp Wave Ripple) という神経活動においては、餌の置かれた目標地点に向けての場所細胞系列が時間的に圧縮された形で再生するリプレイという現象が観測されている。現在位置から離れた場所で起こる場合はリモート・リプレイと呼ばれる(未来志向的)。SWR 活動を意図系列生成と見做せば状態遷移モデル CA3 に対し EC2 からの目標状態(格子表現)の投入により起動されたと考えうる。

一方、動物が移動中にはシータ波の約半周期(70 ミリ秒)に自分の現在位置を含むような形で、場所細胞の系列が賦活する。この現象は位相歳差と呼ばれ、約1/10の時間に圧縮された時系列である。この情報処理が SWR と同様に CA3 の基づくとするならば、これは現在志向的な性質をもつ意図であろう。

## 4.3 学習と実行を分離するシータ波

海馬 CA3 の神経活動状態は、ある時点(ミリ秒オーダ)においては一つの状態しか持ち得ないだろう。しかし行動中の動物における CA3 (状態遷移モデル)は、意図系列を生成しつつ、外界の情報を取り込んだ地図学習の必要がある。

この点を考慮した移動中のマウスの CA3 の動作仮説は以下となる。まずシータ波の一方の半周期において外界の情報を EC2 から取り込むことで再帰結合神経結合に状態遷移の情報を蓄積する学習する。そして反対の半周期において EC2 から目標状態が入力されるとその前の現在状態から目標状態に至る意図系列が生成される。つまりシータリズムを用いて現在状態と目標状態を組み合わせた現在志向的な意図系列を生成する。

<sup>1</sup>強化学習もシミュレータとして状態遷移モデルを用いて学習を加速できるが、タスクに応じてゴール変化させる柔軟さには欠ける。  
<sup>2</sup>ミリ単位の広がりを持つ CA3 の再帰結合により生み出される意図系列の 25msec 潜時の遅いガンマ波は妥当な周波数である。

## 5. トリシナプス回路は汎用な外界認識装置

### 5.1 知的エージェントは如何にして世界を把握するか

エージェントとしての動物が、その知性によって生存率を高めるには、自己に関わる外部世界を出来るだけ正確に把握することが有利だろう。しかし世界のもつ情報の膨大さに比べて、特定のエージェントがもつ貧弱なセンサから得られる有限の入力状態の遷移では、把握できる範囲はごく一部であり、現実的な広い世界を十分にモデル化することはできない。

世界は変化し続けるが部分的に何らかの規則性を持つ(そうした構造がなければ予測は不可能)。よって世界の中で把握しうる範囲を拡大するには、できるだけ単純で時間推移を高精度で予測しうる小さいモデルの集合に分解することが有効である。

たとえばオフィスにおけるエージェントは、壁や床のように変化しない部分については、一旦認識すれば通常はその状態を更新する必要はないし、5分前に部屋に入ってきた同僚が今どこに居るかは確認しなくてもかなりの確度で推定できる。この例はオブジェクトに分解しうるわかりやすい例であるが、テーブル上の本立がテーブルとともに動くような状況まで考慮すれば、本質的にこれがフレーム問題となっていることを理解しうるだろう。

ここで脳について見ると、新皮質と海馬(ここでは嗅内皮質を含めて海馬と呼ぶ)が、外界から入力される状態系列から世界を認識する機能(さらには行動意図を生成する機能)を担う主要な脳器官である。まず時間的に変化しない(少ない)部分に関しては、新皮質がセンサ情報から認識処理を行う各段階においてバッファリングすると考えるのが適当であろう。なぜなら神経回路の場合には同一の表現を複製することはできないので、処理経路上にバッファを持つのが最も自然であるからである。(ここで新皮質上の表現は、センサ入力に直接に設置したものとは限らない、プロ棋士はセンサから切り離された盤面上で操作を行うがその表現は新皮質にあると想定される)

### 5.2 現在志向的意図を取り出す CA1

状態遷移モデルをもち、ナビゲーション機能に重要な役割をはたす海馬が、全体として現在志向的な意図を出力する機能を担うと考えるのは妥当であろう。

動物が移動中の場合には、海馬 CA1 では、CA3 で生成された意図系列(場所表現)の中から、EC3 から得られる現在状態(格子表現された自己位置を含む)からみて次時刻に実現すべき意図を抽出する。興味深いことは、この機能のために EC3 では現在状態を表現する必要性が生じる。海馬に対して新皮質から様々な目標が与えらるだろうから、逆に EC2 への入力目標状態であろう。CA1 にて取り出された意図は、EC 深層に格子表現に変換されてフィードバックされ、同時に Sb では場所表現としてさらなる処理が行われる。

### 5.3 PER/POR と EC 間の座標変換形式

座標変換を行う方法は主に二つある、画像処理をイメージして考える。一つ目は、位置座標をもつ物体や特徴点の集合体として外界を記述し、その位置座標に対して空間変換を行う方法である。この方法は少なくとも工学的には全ての物体に対して同じ変換を行うことが容易でありメトリック自体の変換である。二つ目は、画像データのように空間中に配置された値の行列を想定し、変換前の行列から変換後の行列への対応関係を変更するマルチプレクサによる方法である。局所的な画像の歪みの吸収などは行い易く新皮質を模した深層学習(CNN)のプーリング層で使われている。以下の理由により海馬における座標変換は後者のマルチプレクサであると考えうる。

第一に視覚野で実現されている可能性が高いので神経回路として実装しやすい。第二に、外界のメトリック(ユークリッド空間)を明示的に表現すると目される神経回路である格子細胞が座標変換に関わる MEC に存在しており、物体などの What 情報を扱う LEC には存在しない。部分的にメトリックを明示的に表現する必要があるのは、元の変換がメトリックを含んでいないことを示唆している。第三に、進化的に新皮質より先に先行している海馬の計算機能は汎用性が高いと思われ、そうだとすれば柔軟性の高いマルチプレクサが自然だろう。

#### 5.4 MEC3 での自己位置推定

海馬において自己位置は既に述べたような環境中心座標系における空間選択細胞によって表現される。一方で環境からのあらゆるセンサ入力自己中心座標系であるため、この間での座標変換が必要となる。そこで自己位置推定をおこなうには、環境中心座標系の姿勢情報(位置/方向など)を維持するための機能をトリシナプス回路が有用である。

自己位置推定は Where 経路上の MEC3 にて行われると仮定する。MEC3 は MEC 深層からの予測状態(意図)、外界からの POR を通じた現在入力、ParaSb からの大域的な姿勢情報(現在状態)を入力として受け取る。そして MEC3 中において、現在時刻における環境中心座標系としての位置情報を格子状の活動として表現する。自己位置推定はベイズフィルタ風のアルゴリズムで実行されると考えられ、短時間にはユークリッド空間中で少しの位置しか移動しないという事前知識を利用する。

#### 5.5 予測性を高める DG: パタン分離と等価性構造抽出

前記のように、海馬の推論エンジンとコアとなる CA3 は状態遷移モデルであるため、その入力となる DG において CA3 の時間を含む動作状態を利用して予測性を高める前処理を行う。

機械学習等のデータに基づく帰納推論は、簡単なモデルで内に密にデータが存在するほど予測性能が高い。これを実現するには、高次元空間から適切な部分空間(変数集合の一部、神経科学的にはセル・アセンブリに関連)を選択する手段と、複数の部分空間を結合する等価性構造を発見する手段がある。

以下で、EC2 から歯状回(DG)における顆粒細胞(GC)、さらに GC から CA3 に至る経路における計算機能について述べる。

##### (1) 顆粒細胞が担うパタン分離(部分空間の抽出)

GC にとっての入力となる EC2 における高次元空間から、時系列予測に適した部分空間を選択する機能は、神経科学分野においてパタン分離と呼ばれる機能に一致する。ここでは GC の学習は、以下であると仮定する。

歯状回の部分空間モデル学習仮説: EC2 から CA3 への直接入力を、EC2 から歯状回の顆粒細胞(GC)は DG を経て CA3 への投射する部分時系列パタンの重ね合わせとして置き換えられるように部分空間モデルの学習を行う。

この処理では CA3 から苔状細胞(MC)経由で DG の顆粒細胞(GC)に再帰する神経結合(Inter-item hetero associator)により CA3 における時系列予測の精度に関わる情報をフィードバックすることで、GC のパタン分離の学習を制御する。

##### (2) 新生細胞が支える等価性構造抽出(部分空間の結合)

二つの部分空間の等価性は、そこに含まれる変数を一対一対応させて等価に扱いて得ると定義する。そして相互結合できる等価な部分空間の集合を表現した構造を「等価性構造」と呼び、その抽出技術を等価性構造抽出と呼ぶ。等価性構造に含まれる部分空間は、機能推論において結合できることに実質的にイベント数が増加することで予測性能を高められる。

等価性構造は、画像認識における並進/回転などの不変性(CNN のプリーング層)や、類推における対応関係に関わり、前記のマルチプレクサ型の座標変換の形式を一般化したものでもある。また AI 分野におけるオントロジーマッチングにも関わる。

しかし等価性構造抽出においては探索空間の膨大さが問題となる。変数全体の数を  $N$  とし、部分空間のサイズを  $d$  とすれば、部分空間の組み合わせは  ${}_N C_d$  個である。よって結合する部分空間ペアの組み合わせは  ${}_N C_d ({}_N C_d - 1) d! / 2$  に達する。多様な組み合わせから予測性に関わる評価を行うには、静的なパタンのみでは表現の多様性が不足するため、海馬における一貫した時系列情報に基づく評価が有効である[山川 13]。

等価性構造抽出における処理の本質が、「部分空間の対応付け」であるため、対応を探索する段階では二種類以上に色付けされた部分空間表現を用いつつ、最終的にはそれらをマルチプレクサとして統合する必要がある。部分空間の色付けに GC の新生細胞を利用することで、等価性構造抽出についての次のような処理過程仮説を提案する[井ノ口 13]。

新生顆粒細胞を用いた等価性構造抽出仮説: EC2 入力を CA3 に出力する部分空間を安定的に表現する成熟した GC に対し、新規の経験に対応するイベントを保持した新生 GC が、既存の部分空間との対応を探索する処理として等価性構造を実現する。

上記のようなオンライン学習によれば、成熟した GC によって既に形作られた有用な部分空間の数を  $M_d$  とすれば( $d$  は次元数)、ある時点で新生細胞が探索する範囲は  $M_d \times d!$  という小さな範囲に限定できるため、現実的な計算が可能になる。

等価性構造抽出においては、一貫した時系列処理が必要でなおかつ細胞新生の性質を有用利用できるもので、脳において本処理は海馬 DG においてしか実現し得ないと考えられる。

## 6. おわりに

現状の断片的な神経科学等の知見を繋ぎあわせ、脳を参考とした汎用的な認知アーキテクチャを構築するには、一貫した仮説の集合体としての物語を制作し、それを洗練する研究プロセスが有効である。本稿では海馬と嗅内皮質の計算モデルについての初期的な物語を提案した。ここでは、海馬 CA3 を状態遷移モデルと仮定することを起点に、海馬の中心的なリカレント回路(トリシナプス回路)の計算機能について神経科学知見を参考にモデル化を進めた。今後、この物語をたたき台として、神経科学・AI・計算機科学・認知科学等の視点から改良したい。

本物語が概ね正しいとすると今後検討すべき課題として以下等がある。GC と CA3 周辺で獲得した等価性構造を EC と新皮質間や新皮質内の神経結合に転移させる仕組み。EC2 と EC3 のいずれかが外部から目標状態を受け取るか否かを PER/POR の知見から特定する。時間的な階層性と ParaSb/PreSb との関連の探求。前頭野との連携で実現される作業記憶の機構など。

本稿の執筆にあたり、OIST の五十嵐潤氏、はこだて未来大の佐藤直行氏、さらに理研 BSI の佐藤正晃氏、水田恒太郎氏、イスラム・タンビル氏、山口陽子氏、林康紀氏らに感謝する。

## 参考文献

- [全脳] 全脳アーキ勉強会, <http://www.sig-agi.org/wba>
- [野家 05] 野家啓一, 物語の哲学, 岩波書店 2005.
- [北西 15] 北西卓摩, <http://leading.lifesciencedb.jp/4-e001/>
- [山川 13] 山川宏, JSAI2013, 3H4-OS-05c-2in, 2013.
- [井ノ口 13] 井ノ口馨, 記憶をコントロールする, 2013.
- [Grillner] Handbook of Brain Microcircuits, Oxford, 2010.